Biostatistics 666
Problem Set 6
Due March 30, 2006

Multipoint Analysis

1. Two siblings were genotyped at three consecutive SNP markers. For the first sibling the genotypes A/A, C/C, G/G were observed at the three markers. For the second sibling, genotypes G/G, C/C, G/T were observed at the three markers. Assume that the first marker has alleles {A,G}, the second marker has alleles {C,T} and the last marker has alleles {G,T}. Further, assume that the two alleles at each marker have equal frequencies.

    a) Considering each marker individually, calculate the probability that the two siblings are IBD at each marker location.

    b) Calculate the probability of each possible IBD state for the middle SNP, considering the three markers simultaneously and assuming the recombination fraction between each pair of consecutive markers is $\theta = 0.0528$.

2. Consider two individuals where the first individual has genotype 1/1 at three consecutive loci and the second individual has genotype 2/2 at the same three loci. Assume that alleles 1 and 2 have frequency 0.50 at each locus.

    a) Let the recombination fraction between consecutive loci be $\theta = 0.0528$. Based on the genotype data, what is the relative probability that the two individuals are siblings rather than half-siblings?

    b) Repeat the calculation above assuming the recombination fraction between consecutive loci is $\theta = 0.5$.

    c) Comment on the difference between a) and b). What are the trade-offs that should be considered when deciding between the use of closely spaced markers versus markers that are far apart for relationship inference?

    d) How would you expect your results above to change if your model allowed for genotyping error?